Differentiable Probabilistic Models

William Watson
nextbillyonair@gmail.com

Abstract

d

Contents

1 Introduction **4**

1.1 Philosophy 4

2 Preliminary **4**

2.1 Gradients 4

2.2 Jacobian 4

2.3 Hessian 4

2.4 Newton Optimization 4

3 Distributions **4**

3.1 Distribution 4

3.2 Arcsine 4

3.3 Bernoulli 4

3.4 Beta 4

3.5 Categorical 4

3.6 Cauchy 4

3.7 Chi Square 4

3.8 Conditional Model 4

3.9 Convolution 4

3.10 Data 4

3.11 Dirac Delta 4

3.12 Dirichlet 4

3.13 Exponential 4

3.14 Fisher-Snedcor (F-Distribution) 4

3.15 Gamma 4

3.16 Generator 4

3.17 Gumbel Softmax 4

3.18 Gumbel 4

3.19 Half Cauchy 4

3.20 Half Normal 4

3.21 Hyperbolic Secant 4

3.22 Langevin 4

3.23 Laplace 4

3.24 Log Cauchy 4

3.25 Log Laplace 4

3.26 Log Normal 4

3.27 Logistic 4

3.28 Normal (Multivariate) 4

3.29 Rayleigh 4

3.30 Relaxed Bernoulli . . . 4
3.31 Student T . . . 4
3.32 Transformed Distribution . . . 4
3.33 Uniform . . . 4

4 Mixture Models **4**
4.1 Mixture Model . . . 4
4.2 Gumbel Mixture Model . . . 4
4.3 Infinite Mixture Model . . . 4

5 Transforms **4**
5.1 Transform . . . 4
5.2 Inverse Transform . . . 4
5.3 Affine . . . 4
5.4 Exp . . . 4
5.5 Expm1 . . . 4
5.6 Gumbel . . . 4
5.7 Log . . . 4
5.8 Logit . . . 4
5.9 Power . . . 4
5.10 Reciprocal . . . 4
5.11 Sigmoid . . . 4
5.12 SinhArcsinh . . . 4
5.13 Softplus . . . 4
5.14 Softsign . . . 4
5.15 Square . . . 4
5.16 Tanh . . . 4

6 Divergences **4**
6.1 Cross-Entropy . . . 4
6.2 Perplexity . . . 4
6.3 Forward KL Divergence . . . 4
6.4 Reverse KL Divergence . . . 4
6.5 Jensen-Shannon Divergence . . . 4
6.6 Earth Mover's Distance . . . 4

7 ELBO **4**

8 Models **4**
8.1 Linear Model . . . 4
8.2 Linear Regression (Normal) . . . 4

8.3 L1 Regression (Laplace) 4
8.4 Ridge Regression (Normal + Normal Prior on Weights) 4
8.5 Lasso Regression (Normal + Laplace Prior on Weights) 4
8.6 Logistic Regression (Bernoulli) 4
8.7 Softmax Regression (Categorical) 4

9 Adversarial Loss **4**

9.1 Adversarial Loss 4
9.2 GAN Loss 4
9.3 MMGAN Loss 4
9.4 WGAN Loss 4
9.5 LSGAN Loss 4
9.6 Gradient Penalty 4
9.7 Spectral Norm 4

10 Monte Carlo **4**

10.1 Monte Carlo Approximation 4
10.2 Linear Congruential Generator 4
10.3 Inverse Transform Sampling 4
10.4 Box-Muller 4
10.5 Marsaglia-Bray 4
10.6 Rejection Sampling 4

11 Markov Chain Monte Carlo (MCMC) **4**

11.1 Metropolis 4
11.2 Metropolis-Hastings 4
11.3 Metropolis-Adjusted Langevin Algorithm (MALA) 4
11.4 Hamiltonian Monte Carlo 4

1 Introduction

1.1 Philosophy

2 Preliminary

2.1 Kronecker Product

2.2 Gradients

2.3 Jacobian

2.4 Hessian

2.5 Newton Optimization

3 Distributions

3.1 Distribution

3.2 Arcsine

3.3 Bernoulli

3.4 Beta

3.5 Categorical

3.6 Cauchy

3.7 Chi Square

3.8 Conditional Model

3.9 Convolution

3.10 Data

3.11 Dirac Delta

3.12 Dirichlet

3.13 Exponential

3.14 Fisher-Snedcor (F-Distribution)

3.15 Gamma

3.16 Generator

3.17 Gumbel Softmax

3.18 Gumbel

3.19 Half Cauchy

3.20 Half Normal

3.21 Hyperbolic Secant

3.22 Langevin

3.23 Laplace

3.24 Log Cauchy

3.25 Log Laplace

3.26 Log Normal

3.27 Logistic